

Methodology article

Iterative Group Analysis (iGA): A simple tool to enhance sensitivity and facilitate interpretation of microarray experiments

Rainer Breitling^{*1,2}, Anna Amtmann¹ and Pawel Herzyk^{2,3}Address: ¹Plant Science Group, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom,²Bioinformatics Research Centre, Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, United Kingdom and ³Sir Henry Wellcome Functional Genomics Facility, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United KingdomEmail: Rainer Breitling^{*} - r.breitling@bio.gla.ac.uk; Anna Amtmann - a.amtmann@bio.gla.ac.uk; Pawel Herzyk - p.herzyk@bio.gla.ac.uk^{*} Corresponding author

Published: 29 March 2004

Received: 15 November 2003

BMC Bioinformatics 2004, **5**:34

Accepted: 29 March 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/34>

© 2004 Breitling et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The biological interpretation of even a simple microarray experiment can be a challenging and highly complex task. Here we present a new method (Iterative Group Analysis) to facilitate, improve, and accelerate this process.

Results: Our Iterative Group Analysis approach (iGA) uses elementary statistics to identify those functional classes of genes that are significantly changed in an experiment and at the same time determines which of the class members are most likely to be differentially expressed. iGA does not require that all members of a class change and is therefore robust against imperfect class assignments, which can be derived from public sources (e.g. GeneOntologies) or automated processes (e.g. key word extraction from gene names).

In contrast to previous non-iterative approaches, iGA does not depend on the availability of fixed lists of differentially expressed genes, and thus can be used to increase the sensitivity of gene detection especially in very noisy or small data sets. In the extreme, iGA can even produce statistically meaningful results without any experimental replication.

The automated functional annotation provided by iGA greatly reduces the complexity of microarray results and facilitates the interpretation process. In addition, iGA can be used as a fast and efficient tool for the platform-independent comparison of a microarray experiment to the vast number of published results, automatically highlighting shared genes of potential interest.

Conclusions: By applying iGA to a wide variety of data from diverse organisms and platforms we show that this approach enhances and accelerates the interpretation of microarray experiments.

Background

Microarray experiments determine the relative expression levels of a large number of genes in various conditions. In the easiest case they are used to detect differences in gene expression between two conditions, e.g. in diseased vs. healthy tissue, or in mutant vs. wild type organisms. The results in general are presented as lists of "differentially

expressed" genes. A number of statistical techniques have been successfully employed to prepare and analyze these lists (preparation reviewed in [1-5]; analysis methods e.g. [6-11]).

In some applications, e.g. identification of marker genes or potential drug targets, such mere lists of genes may be

considered sufficient. More often, however, it will be necessary to make biological sense of the data. Given lists of hundreds of genes, this is a daunting task, especially when nothing is known about the expected outcome. But even when the physiology of the experimental system is well-defined, it is sometimes difficult to distinguish which parts of the interpretation are significant and which may be artifacts of a biased focus on familiar features.

Here we present a method (Iterative Group Analysis, iGA) that provides an automatic functional annotation of microarray results together with a statistical confidence level for each annotation feature. This annotation is based on a comprehensive hypergeometric statistics calculation detecting concerted changes in "functional classes" of genes. In contrast to previous methods with similar objectives [6-11], iGA does not require a fixed list of reliably differentially expressed genes but rather uses an iterative procedure to determine the optimal threshold for each functional class. This feature should make iGA more flexible in the case of noisy data, when a reliable list of genes is not available, and allows its use as a sensitive gene detection method. The functional classes can be of diverse origin (e.g. GeneOntology assignments <http://www.geneontology.org>, BLAST result key words, literature extracts) and the detection algorithm will automatically determine the genes in each class that are most likely to be differentially expressed.

By focusing on groups instead of single genes, it is possible to determine statistical significance even without experimental replication, the group members serving as "internal replicates". While this will not provide the same detail as "real" replication, it can provide important information on the biology of a sample. At the same time, the iGA approach should enhance the sensitivity of gene detection, especially for small, noisy data sets.

The iGA approach also provides a fast and efficient way to compare an experiment with a large number of published microarray experiments, without requiring a common experimental platform or analytical technique.

Results and discussion

Functional annotation of microarray results

Iterative Group Analysis is based on the idea that a concerted expression change of (some) members of the same functional class is physiologically relevant. We assign each gene to functional classes, e.g. based on its GeneOntology assignments. For each functional class we then determine the probability of change (PC-value). The iterative procedure for that calculation is illustrated in Fig. 1. First, all genes are sorted according to a metrics of differential expression (fold-change, *t*-statistics...). The choice of the optimal sorting method will depend on personal prefer-

ences and details of the experimental design, but has relatively little influence on the iGA results (not shown; see also [8]). In the test cases described below we usually sorted the genes once by up-regulation and once by down-regulation and analyzed the two lists separately, but sorting by absolute changes is possible just as well. Second, we move along the list, counting the members of the functional class of interest, and each time we encounter a new class member we ask: How likely is it to observe this many members of the class that high up in the list by chance. This probability (*p*-value) is exactly:

$$p(z, n, t, x) = p(z-1, n, t, x) - \frac{\binom{t}{z} \binom{n-t}{x-z}}{\binom{n}{x}}; p(0, n, t, x) = 1,$$

where *n* is the total number of genes, *x* is the total number of class members, and *t* is the rank of the *z*-th class member (*z* being the current step in the iteration; see Fig. 1).

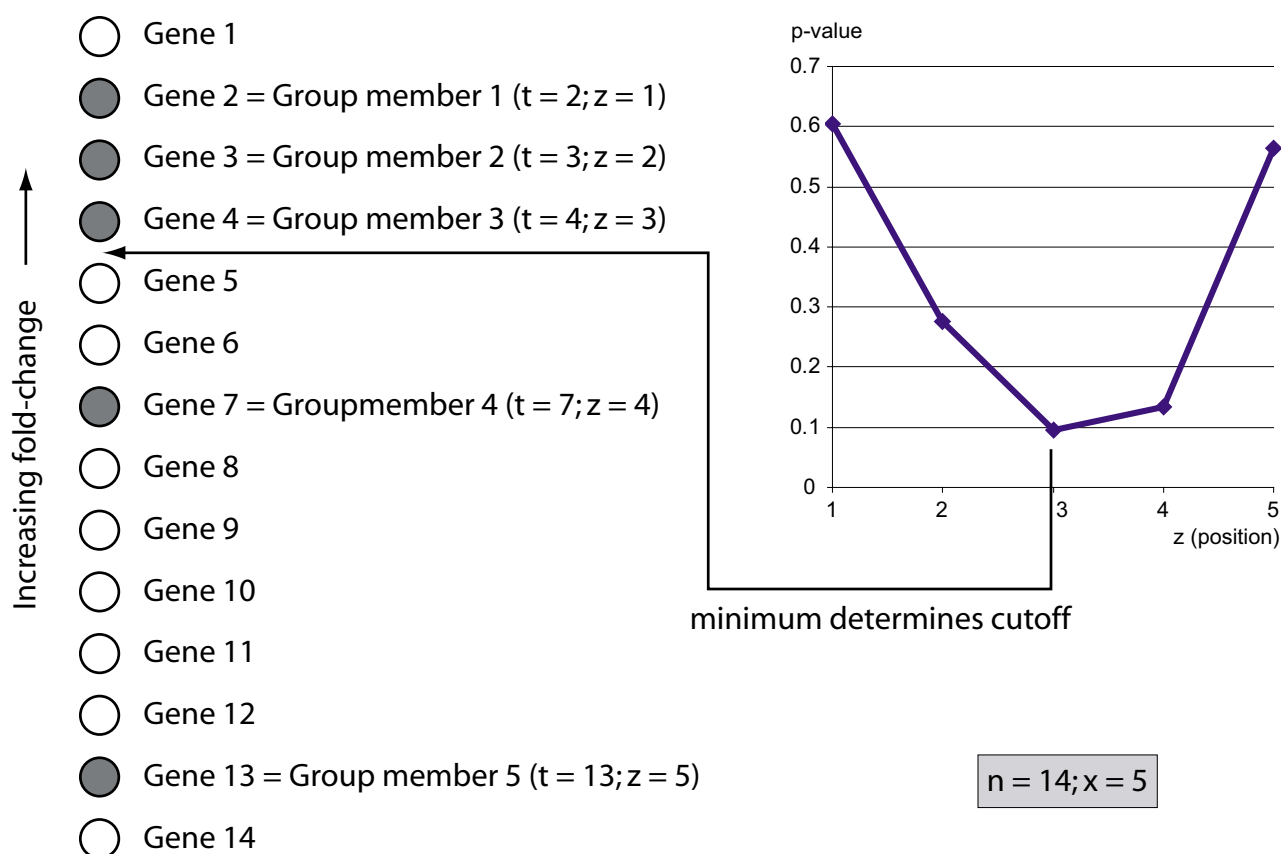
The notation $\binom{n}{k}$ indicates the binomial coefficient, i.e.

the number of ways of picking *k* unordered items from a list of *n* items. This equation is equivalent to that used by Onto-Express [7] in a more restricted context. Third, we determine the position in the list that yields the smallest *p*-value, and assign this as the PC-value of the class. All class members above that position are considered as "potentially differentially expressed". Fourth, all classes are sorted by their PC-values. The classes with the lowest PC-value are most significantly changed.

Note that this process does not require that all members of a group change at the same time or in the same direction – a very important feature, because genes that share a functional annotation may include activators as well as inhibitors of a certain process, and may also include a large number of hypothetically assigned genes that in the majority won't change at all.

In the case of genes that are present several times on a chip, the user can decide how to proceed: Either the replicate spots are merged into an average value during the previous steps of analysis, i.e. before sorting the gene list, or each replicate is treated as an independent group member. The latter approach is statistically not quite correct, as replicate spots do not represent independent measurements, but in the case studies described below we found the resulting summaries of replicates very useful.

An example illustrating the use of iGA is shown in Table 1. In this experiment, *Arabidopsis thaliana* seedlings were grown in the dark for 4 days and then treated with white

**Figure 1**

Principle of Iterative Group Analysis. The left panels shows a notional microarray result for 14 genes ($n = 14$), which are sorted by decreasing fold-change. 5 of the genes (filled circles) belong to the functional class of interest ($x = 5$). For each class member the p -value was calculated according to the hypergeometric equation given in the text, using the t - and z -values shown next to each gene. The left panel shows those p -values plotted against the position of the class member. The minimum is found at position 3 and is used to determine the cutoff for this group, i.e. group members 1 to 3 would be listed as "most likely to be up-regulated". The corresponding p -value (0.1) would be assigned as the PC-value for this group.

light for six hours (chip ID 7341, see Material and Methods). The dark-grown seedlings are almost colorless, because they lack the photosynthetic pigments, and after light treatment the most obvious physiological reaction of the plants is their de-etiolation ("greening"), i.e. they rebuild the photosynthetic machinery. Gene expression was compared to non-treated plants on a single microarray. 460 of 4828 detectable genes are more than 2-fold up-regulated under these conditions. All genes were sorted by fold-change and 588 GeneOntology classes were tested for differential expression. Table 1 shows that iGA reliably identifies the effect on the photosynthetic apparatus (6 out of the top 10 groups) and also detects changes in biochemical pathways that may be associated with the restart

of anabolic metabolism (pentose-phosphate cycle, starch breakdown).

Assigning statistical confidence to the annotation

As the PC-values are directly derived from p -values, they already give a good idea of the statistical significance of an observed change, after correcting for the effect of multiple testing, i.e. the fact that many hundreds of groups are tested for differential expression at the same time. However, as many functional classes overlap, multiple testing correction using the total number of functional classes tested is too restrictive. At the same time, the PC-values are underestimating the true probability of changes, because they are based on determining the minimum p -value within each class. As iGA is rank-based, it is very easy to

Table 1: Up-regulated GeneOntology classes after 6 h of white light-treatment applied to etiolated *A. thaliana* seedlings (chip ID 7341, see Material and Methods for details). The top 10 classes are shown. They contain a total of 26 distinct genes that are detected as "possibly changed". Six of the 10 classes are directly related to the light reactions of photosynthesis.

GO class	GO class description	PC-value	# changed	# members
GO:0015979	Photosynthesis	2.99E-08	5	8
GO:0009538	photosystem I reaction center	4.52E-07	4	5
GO:0019253	reductive pentose-phosphate cycle	7.27E-05	3	12
GO:0016168	chlorophyll binding	9.40E-05	3	6
GO:0009523	photosystem II	2.36E-04	2	2
GO:0000311	plastid large ribosomal subunit	9.26E-04	3	8
GO:0016459	Myosin	1.75E-03	1	6
GO:0004556	alpha-amylase activity	1.94E-03	3	3
GO:0003774	motor activity	2.05E-03	1	7
GO:0009533	chloroplast stromal thylakoid	2.14E-03	2	2

overcome these two problems by calculating PC-values from a large number of random permutations of genes and count how often a given p -value and smaller is actually observed in those random data. This approach is possible even when the list of genes is based on a single experiment. E.g., PC-values smaller than 0.00092 occurred 48 times in 100 random permutations of the photosynthesis example discussed above. This also means that the expected false-discovery rate (FDR) among the top 5 groups in the example is less than 10% [12]. Such an FDR is not particularly impressive at first glance, but these results are based on a single array, while in a single-gene approach no statistical confidence determination is possible at all without replication (or a detailed *a priori* knowledge of measurement errors). The results can be corroborated by the analysis of three replicate hybridizations for the same experiment, as iGA finds almost exactly the same groups in two of them (the third [chip 14753] may be a biological "outlier", as its iGA result shows hardly any similarity to the other three chips, although clustering analysis does not indicate a serious technical problem with this hybridization).

Major assumptions underlying the iGA statistics are the independence of measurements within a group and equal variation between groups. While the first assumption will generally hold unless replicate spots or cross-hybridizing isoforms of a gene are included as group members, the second assumption may be violated for biological reasons: There may be certain groups of genes that tend to show more variation of their expression than others, even in constant conditions, and will therefore show up more often in the iGA results. The same problem occurs with a manual analysis of microarrays. iGA is still applicable in these cases, but is essential that the statistical analysis is followed by a rigorous assessment of the biological significance of the results.

Sources of functional annotations

The power of iGA increases with the quality of the available annotations. In the case of human and mouse genomes, the functional annotation by GeneOntology terms is relatively detailed and comprehensive and in our test cases (see below) gave satisfactory results. For rat and *A. thaliana* data we found it useful to include classes defined by key words automatically extracted from the complete annotation of each gene (including the gene description and the description of BLAST hits). While those key word-based classes are not necessarily biologically meaningful, iGA proved to be sufficiently robust to detect the most relevant changes. Of course, user-defined groups of interest can be particularly powerful and are easily integrated in iGA.

Sensitive gene detection by using Iterative Group Analysis

Instead of focusing on a list of "differentially expressed" genes that is arbitrarily determined by some statistical threshold, we make use of all genes, sorted by their level of "differential expression". This provides the additional opportunity of using iGA as a sensitive tool to detect differentially expressed genes, if they are part of a group of genes changing in concert. Such cases can be biologically very interesting. E.g., when a certain metabolic pathway is changed, only a few rate-controlling enzymes may change dramatically and those could even be absent on the particular microarray used. The rest of the pathway components will change only slightly, but if all of them move in the same direction, this important effect will be detected by iGA. Another example is the detection of tissue-specific regulation in the thylakoids, even in the sentence: "In a study of gene expression in roots of *A. thaliana* seedlings during potassium starvation we expected important changes in root potassium transporters. However, if each cell type along the way from soil to internal vessel would specifically change the expression of only a few transport-

Table 2: Group analysis of gene expression in potassium-starved *A. thaliana* roots. This experiment used a custom-made transporter array and corresponding annotations [13]. The PC-values, number of group members and of significantly changed genes are indicated. Only groups with a group-wise FDR <10% are shown. The last column indicates the number of group members that were detected by Significance Analysis of Microarrays [14] with a FDR <10%.

Group	PC-value	Group members	Changed members	SAM
Potassium transporter (up)	1.26E-05	13	7	2
Nitrate transporter (down)	3.47E-04	9	3	1
Putative anion exchanger (up)	8.62E-04	7	2	2
19 TMS proteins (down)	2.22E-03	2	2	0
			Total: 14	Total: 55

Table 3: Comparison of SAM and iGA performance on noisy lymphocyte data. The cells were obtained from elicitor-treated animals and show an expression pattern indicative of immune system activation. For iGA the genes on the array were annotated by Gene Ontology terms and keywords extracted from gene names.

	SAM	iGA (based on lists of sorted by fold-change)
Groups detected	n.a.	Kallikrein (4 genes) Rhesus transporters (2) Antimicrobial peptides (4) Classical complement (4) Immunoglobulins (6) Scavenger receptors (2) 7 other groups (15)
Genes detected	0	37 total
FDR used	50%	10%

ers, these expression changes would be "diluted" in the whole-root analysis. In fact, a single-gene approach found only two significantly up-regulated potassium transporters among 55 genes that were labeled as "differentially expressed". In contrast, iGA identified potassium transporters as the most significantly changed group of all in this experiment, with 7 out of 13 potassium transporters involved in the response (3 replicates of a membrane transporter microarray, [13]; Tab. 2).

The focus on functionally consistent gene groups can be particularly helpful in very noisy data sets, where the results lists are highly contaminated by false positives. In one of the test cases described in the next section, each sample consisted of a purified lymphocyte subpopulation isolated from single mice. Due to inter-individual variation and the tiny amount of material the data were so noisy that Significance Analysis of Microarrays [14] failed to find any differentially expressed genes with a false discovery rate below 50%. iGA analysis of fold-change-based lists in contrast found 37 genes in about 10 functional classes and made biological sense of the data (Tab. 3).

Using Iterative Group Analysis to compare microarray results

A variation of iGA uses classes based on the results of previous microarray experiments. Those classes can either be manually extracted from the published literature (e.g., each list of up- or down-regulated genes defines one class) or automatically determined from the raw data files (e.g., each class contains the genes that change more than 2-fold in a certain experiment). In this way, each experiment is described by a number of "signature classes", which can then be tested for concerted differential expression in the experiment of interest. This approach not only identifies the experimental conditions that come closest to the present experiment, but also highlights the shared genes. In contrast to clustering techniques that can be used for similar purposes, iGA does not require that the same genes are present in all experiments, that the experimental platform is the same, or that all genes behave identically. The "open", iterative technique of iGA makes it more robust than previous "fixed list"-based approaches used for similar purposes [15]. It is also not necessary that the complete data are available for all experiments. The statistical confidence measures associated with iGA results are particularly useful when only a small number of regulated

Table 4: Summary of the case studies used for the "semi-blind" evaluation of iGA. The first two columns indicate the study organism and experiment performed. The next two columns show the tissue and physiological condition identified on the basis of iGA.

Organism	Experimental condition	Tissue identified	Condition identified	Remarks
Mouse	activated T-cells	leukocytes	infection	
Human	CaCo-2 cell Campylobacter infection	cell culture	anoxia, growth arrest, immediate-early response	
Human	CaCo-2 cell Campylobacter infection	cell culture – hepatocytes	growth arrest, acute-phase response	1
Human	polymyositis biopsies	muscle biopsies	inflammation, anaerobiosis	
Human	Chronic Fatigue Syndrome muscle	muscle biopsies	"older muscle"	
Human	lipid phosphate phosphatase-transfected HEK cells	embryonal neurons	mostly minor changes	2
Mouse	testis of hormone receptor knock-outs	spermatogenic tissue	steroidogenesis vs. spermatogenesis	3
Rat	phaeochromocytoma treated with neural growth factor	neurons?	growth-factor response vs. basic metabolism	4
Human	activated monocytes	myeloid leukocytes	inflammation; early activation; chemokines	

1 CaCo2 cells have an hepatocyte-like acute-phase response [16] 2 HEK cells have important neuronal features [17] 3 differences in sperm number are the main effect of the gene disruptions 4 phaeochromocytoma are neural-crest derived cells

genes are shared between two experiments, as we often found to be the case even for almost identical treatments (not shown).

Validation of Iterative Group Analysis

To validate our approach we performed a "semi-blind" study of microarray data produced at the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow. The data were collected by one of us (P.H.) from samples provided by several collaborators. The pre-processed and normalized data were then passed on to iGA without information on the biology of the experiment other than the array used. Using these "anonymized" results the first author (R.B.) produced a detailed description of the samples, trying to identify the sample source, treatment and physiological status of each sample. The final interpretation was then discussed with the experimentalists to establish the correctness of the iGA conclusions.

Overall, the results of the "semi-blind" study showed that iGA provides a detailed and largely correct picture of the biology of the samples (Tab. 4). In 6 out of 9 cases the tissue was identified correctly, in 7 out of 9 cases the interpretation of the physiological response came very close to identifying the actual treatment, sometimes approaching what a collaborating experimentalist called a "comprehensive molecular diagnosis" of their sample. In no case did iGA miss changes that were considered relevant by the experimentalist after a comprehensive manual survey. Not all experiments were equally amenable to interpretation, either manually based on gene lists or using iGA, however, it was noticeable that in each case iGA speeded up the interpretation process drastically (anecdotally

speaking, by a factor of about 10, although the exact improvement will depend on the analyst and the specific experiment). This was all the more striking as the analysis did not benefit from the extensive background of expectations and experiment-specific expertise that normally guides the analysis process.

Conclusions

This study shows that iGA can yield a reasonable, unbiased, and statistically supported interpretation of microarray data even without prior knowledge of the expected outcome. With increasing comprehensiveness and quality of available annotations the method will become ever more powerful. iGA can be used as a stand-alone application, but it should also be easily integrated with existing graphical interfaces for microarray exploration such as MAPPFinder [6], SuperViewer [10], GoMiner [11], or various commercial programs.

Material and methods

Pre-processed microarray data for light-treated *A. thaliana* seedlings were obtained from TAIR (<http://www.arabidopsis.org>; ExpressionSet:1005823603, AFGC IDs 7341, 14753, 14765, 14779). These were from two-color cDNA arrays that covered about 5000 genes after pre-processing as described on the Arabidopsis website. All other experiments were performed at the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow <http://www.gla.ac.uk/functionalgenomics/> and mainly involved whole-genome Affymetrix arrays for various mammalian species. Details of those experiments will be published elsewhere. Mammalian gene annotations were downloaded from Affymetrix <http://www.affyme>

trix.com/analysis/download_center.affx, *A. thaliana* gene annotations were obtained from TAIR.

iGA was implemented as a Perl script (Additional file 2). Required input data are a list of genes sorted by differential expression (Additional file 4) and a list of functional annotations (Additional file 5). A Windows executable of iGA (Additional file 1) as well as a manual (Additional file 3) and further annotation and examples (Additional files 6, 7, and 8) are also provided.

Authors' contributions

RB devised, implemented and tested the iGA approach and drafted the manuscript. PH co-designed the "semi-blind" test and collected data for it. PH and AA supervised the project. All authors read and approved the final manuscript.

Additional material

Additional File 1

iGA software. For use from the Windows command line.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S1.exe>]

Additional File 2

iGA source code.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S2.pl>]

Additional File 3

iGA Manual. This file describes the use of the iGA software (both the Windows and Perl version).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S3.pdf>]

Additional File 4

Example input file for use with iGA. This file contains a sorted list of genes from a photosynthesis experiment similar to those described in the text.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S4.txt>]

Additional File 5

Gene annotations file for use with iGA. This file contains a classification of genes in the example input file by GeneOntology numbers.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S5.txt>]

Additional File 6

Gene names file for use with iGA. This file contains the descriptive names of the genes in the example input file.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S6.txt>]

Additional File 7

Group names file for use with iGA. This file links GeneOntology terms to GeneOntology numbers.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S7.txt>]

Additional File 8

Sample output file. This file was produced by running iGA with default settings on the four input files (Additional files 4 to 7).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-5-34-S8.txt>]

Acknowledgements

We thank Patrick Armengaud, Wilhelmina Behan, Simone Boldt, Anna Casburn-Jones, Gillian Douce, Paul Everest, Michael Farthing, Heather Johnston, Walter Kolch, Peter O'Shaughnessy, Susan Pyne, Rosemary Smith, and Hawys Williams, who allowed us to test iGA on their data and were always willing to discuss their results. This work was supported by BBSRC grants 17/GG17989 (AA and PH) and 17/PI7237 (AA).

References

1. Cui X, Churchill GA: **Statistical tests for differential expression in cDNA microarray experiments.** *Genome Biol* 2003, **4**:210.
2. Pan W: **A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments.** *Bioinformatics* 2002, **18**:546-554.
3. Pan W: **On the use of permutation in and the performance of a class of nonparametric methods to detect differential gene expression.** *Bioinformatics* 2003, **19**:1333-1340.
4. Pepe MS, Longton G, Anderson GL, Schummer M: **Selecting differentially expressed genes from microarray experiments.** *Biometrics* 2003, **59**:133-142.
5. Stolovitzky G: **Gene selection in microarray data: the elephant, the blind men and our algorithms.** *Curr Opin Struct Biol* 2003, **13**:370-376.
6. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR: **MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data.** *Genome Biol* 2003, **4**:R7.
7. Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA: **Global functional profiling of gene expression.** *Genomics* 2003, **81**:98-104.
8. Hosack DA, Dennis G., Jr., Sherman BT, Lane HC, Lempicki RA: **Identifying biological themes within lists of genes with EASE.** *Genome Biol* 2003, **4**:R70.
9. Kim CC, Falkow S: **Significance analysis of lexical bias in microarray data.** *BMC Bioinformatics* 2003, **4**:12.
10. Provart NJ, Zhu T: **A Browser-based Functional Classification SuperViewer for Arabidopsis Genomics.** *Currents in Computational Molecular Biology* 2003, **2003**:271-272.
11. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN: **GoMiner: a resource for biological interpretation of genomic and proteomic data.** *Genome Biol* 2003, **4**:R28.
12. Storey JD: **The positive false discovery rate: A Bayesian interpretation and the q-value.** *Ann. Stat.* in press.
13. Maathuis FJ, Filatov V, Herzyk P, Krijger GC, Axelsen KB, Chen S, Green BJ, Li Y, Madagan KL, Sanchez-Fernandez R, Forde BG, Palmgren MG, Rea PA, Williams LE, Sanders D, Amtmann A: **Transcriptome analysis of root transporters reveals participation of multiple gene families in the response to cation stress.** *Plant J* 2003, **35**:675-692.

14. Tusher VG, Tibshirani R, Chu G: **Significance analysis of micro-arrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98**:5116-5121.
15. Smid M, Dorssers LC, Jenster G: **Venn Mapping: clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes.** *Bioinformatics* 2003, **19**:2065-2071.
16. Molmenti EP, Ziambaras T, Perlmutter DH: **Evidence for an acute phase response in human intestinal epithelial cells.** *J Biol Chem* 1993, **268**:14116-14124.
17. Shaw G, Morse S, Ararat M, Graham FL: **Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells.** *Faseb J* 2002, **16**:869-871.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

